

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of

ANDREAS KELLNER ET AL

Attty. Docket

PHDE 000126

Serial No.:

Filed: CONCURRENTLY

Title: METHOD OF CONTROLLING DEVICES VIA SPEECH SIGNALS, MORE PARTICULARLY, IN MOTORCARS

Commissioner for Patents  
Washington, D.C. 20231


AUTHORIZATION PURSUANT TO 37 CFR §1.136(a)(3)  
AND TO CHARGE DEPOSIT ACCOUNT

Sir:

The Commissioner is hereby requested and authorized to treat any concurrent or future reply in this application requiring a petition for extension of time for its timely submission, as incorporating a petition for extension of time for the appropriate length of time.

Please charge any additional fees which may now or in the future be required in this application, including extension of time fees, but excluding the issue fee unless explicitly requested to do so, and credit any overpayment, to Deposit Account No. 14-1270.

Respectfully submitted,

By   
Edward W. Goodman, Reg. 28,613  
Attorney  
(914) 333-9611

#4 D. J. 10-20-02  
Priority Papers  
jc996 U.S. PTO  
09/935235  
08/22/01



1c996 U.S. PRO

09/935235



08/22/01

**Prioritätsbescheinigung über die Einreichung  
einer Patentanmeldung**

**Aktenzeichen:** 100 41 456.7

**Anmeldetag:** 23. August 2000

**Anmelder/Inhaber:** Philips Corporate Intellectual Property GmbH,  
Hamburg/DE

**Bezeichnung:** Verfahren zum Steuern von Geräten mittels  
Sprachsignalen, insbesondere bei Kraftfahrzeugen

**IPC:** G 10 L, B 60 R

**Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ur-  
sprünglichen Unterlagen dieser Patentanmeldung.**

München, den 10. Mai 2001  
**Deutsches Patent- und Markenamt**  
**Der Präsident**  
Im Auftrag

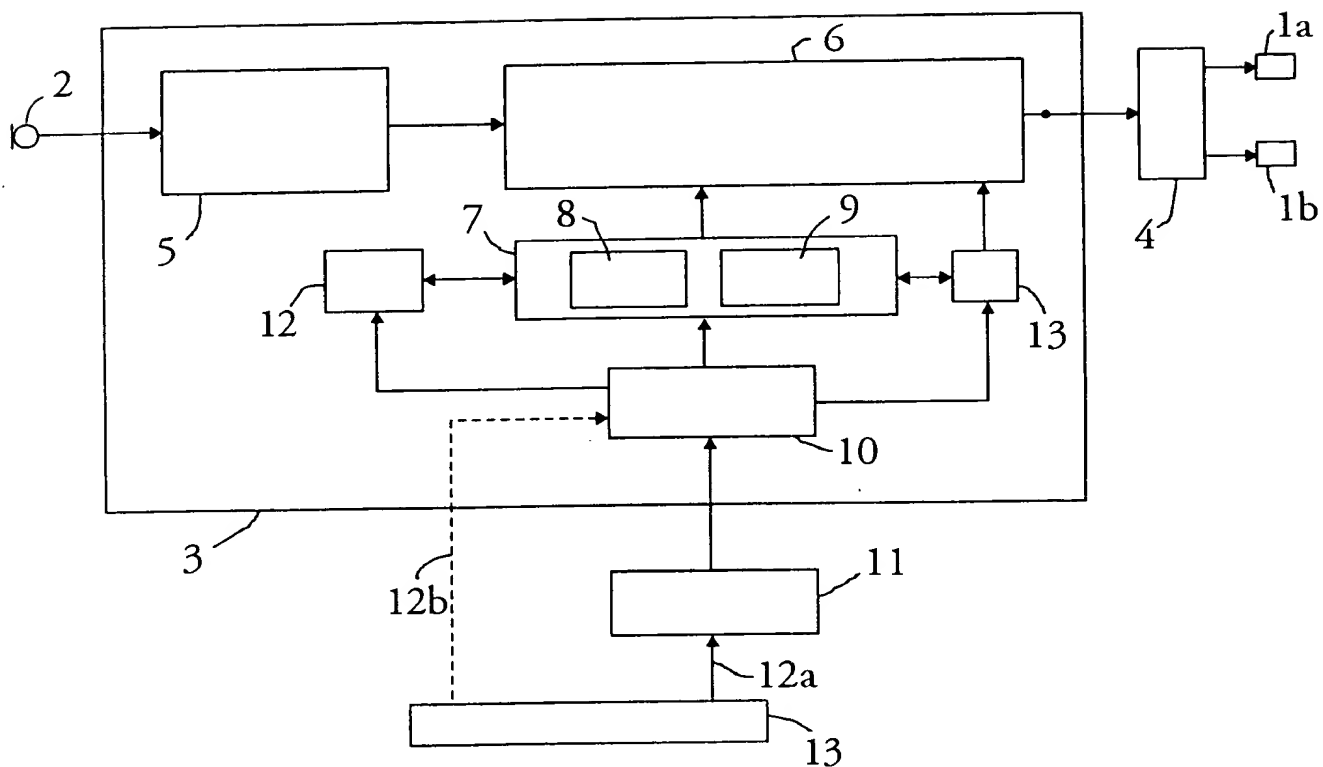
**Dzierzon**

ZUSAMMENFASSUNG

Verfahren zum Steuern von Geräten mittels Sprachsignalen, insbesondere bei Kraftfahrzeugen

- Die Erfindung betrifft ein Verfahren zum Steuern von Funktionseinheiten eines Kraftfahr-  
5 zeugs oder von in einem Kraftfahrzeug angeordneten Geräten (1a, 1b) mittels Sprach-  
signalen, bei dem
- im Kraftfahrzeug auftretende akustische Signale, die vom Betriebszustand und/oder der  
Betriebsumgebung des Kraftfahrzeugs abhängige Störsignalanteile und gegebenenfalls  
Sprachsignalanteile enthalten, einem Spracherkennungssystem (3) zugeführt werden  
10 und
  - das Spracherkennungssystem (3) akustische Referenzen (8) verwendet, die in  
Abhängigkeit von detektierten Betriebszustands- und/oder Betriebsumgebungsinforma-  
tionen ausgewählt und/oder adaptiert werden.
- 15 Die Erfindung ist nicht auf die Sprachsteuerung im Bereich Kraftfahrzeuge beschränkt.

Fig. 1



## BESCHREIBUNG

Verfahren zum Steuern von Geräten mittels Sprachsignalen, insbesondere bei Kraftfahrzeugen

Die Erfindung betrifft ein Verfahren zum Steuern von Funktionseinheiten eines Kraftfahrzeugs oder von in einem Kraftfahrzeug angeordneten Geräten mittels Sprachsignalen. Die Erfindung betrifft auch eine Geräteanordnung zur Durchführung dieses Verfahrens.

Grundsätzlich lässt sich der erfindungsgemäße Ansatz, bei beliebigen Geräten mit einer Sprachsteuerung anwenden, bei denen dem Gerät zugeführte Störsignalanteile vom Betriebszustand und/oder der Betriebsumgebung des jeweiligen Geräts abhängen.

Beim Steuern von Funktionseinheiten eines Kraftfahrzeugs (beispielsweise Steuern eines Scheibenwischerantriebs) und von in einem Kraftfahrzeug angeordneten Geräten (beispielsweise Steuern eines Radios, eines Navigationssystems oder eines Mobiltelefons) mittels von einem Spracherkennungssystem zu erkennenden Sprachsignalen müssen Störsignale berücksichtigt werden, die vom Betriebszustand und/oder der Betriebsumgebung des Kraftfahrzeugs abhängen, um eine fehlerhafte Steuerung der Funktionseinheiten bzw. Geräte zu vermeiden.

Aus der JP 57-30913 (A) ist es bekannt, sowohl die Geschwindigkeit eines Kraftfahrzeugs als auch den eingelegten Gang mittels Sensoren zu detektieren. Aus den Sensorsignalen wird eine Rauschsignalreferenzspannung erzeugt, die ein Maß für den aktuellen Rauschpegel (Störsignalpegel) im Kraftfahrzeug angibt. Die Rauschsignalreferenzspannung wird mit der Ausgangsspannung einer Spracheingabeeinheit verglichen. Beim Vorhandensein von Sprachsteuersignalen empfängt die Spracheingabeeinheit akustische Signale, die sowohl Störsignalanteile als auch Sprachsignalanteile enthalten, was sich in der Ausgangsspannung der Spracheingabeeinheit widerspiegelt. Die Ausgangsspannung der Spracheingabeeinheit wird mit der Rauschsignalreferenzspannung verglichen. Falls die Ausgangsspannung der Spracheingabeeinheit größer als die Rauschsignalreferenzspannung ist, wird ein Spracherkennungssystem aktiviert. Falls die Ausgangsspannung der Spracheingabeeinheit unter die

Rauschsignalreferenzspannung abfällt, wird das Spracherkennungssystem deaktiviert.

Aus der JP 6-83387 (A) ist es bekannt, bei einem Kraftfahrzeug einen Vibrationssensor vorzusehen, um das Vibrieren des Kraftfahrzeugs als Rauschquelle abzuschätzen. Im

- 5 Kraftfahrzeug ist außerdem ein erstes Mikrofon angeordnet, um im Kraftfahrzeuginnenraum vorhandene Störsignale zu detektieren. Ein zweites Mikrofon im Kraftfahrzeuginnenraum dient zur Detektion von Sprachsignalen, die mit Hilfe eines Spracherkennungssystems erkannt werden sollen. Das zweite Mikrofon empfängt allerdings akustische Signale, die neben Sprachsignalanteilen auch Störsignalanteile enthalten. Mit Hilfe der
- 10 Signale des Vibrationssensors, der Mikrofondesignale des ersten Mikrofons und zweier adaptiver Filter wird der Störsignalpegel in den vom zweiten Mikrofon erzeugten Mikrofondesignalen reduziert; die so erzeugten Signale mit reduzierten Störsignalanteilen werden einem Spracherkennungssystem zugeführt.

- 15 Der Erfindung liegt die Aufgabe zugrunde, beim eingangs genannten Verfahren Störsignaleinflüssen wirkungsvoll entgegenzuwirken.

Die Aufgabe wird für Kraftfahrzeuganwendungen dadurch gelöst, dass

- im Kraftfahrzeug auftretende akustische Signale, die vom Betriebszustand des
- 20 Kraftfahrzeugs abhängige Störsignalanteile und gegebenenfalls Sprachsignalanteile enthalten, einem Spracherkennungssystem zugeführt werden und
- das Spracherkennungssystem akustische Referenzen verwendet, die in Abhängigkeit von detektierten Betriebszustands- und/oder Betriebsumgebungsinformationen ausgewählt und/oder adaptiert werden.

25

- Das erfindungsgemäße Verfahren hat den Vorteil, dass auf der Basis häufig leicht zu ermittelnder Betriebszustands- oder Betriebsumgebungsinformationen eine geeignete Anpassung des Satzes für die automatische Spracherkennung zu verwendender akustischer Referenzen erfolgt. Bei einem Kraftfahrzeug können Betriebszustands- oder Betriebsum-
- 30 gebungsinformationen beispielsweise aus einem Bordcomputer ausgelesen werden, der mit ein oder mehreren Detektoren zur Ermittlung des Betriebszustands oder der Betriebsumgebung des Kraftfahrzeugs verbunden ist. Ausgehend vom ermittelten Betriebszustand

bzw. der ermittelten Betriebsumgebung werden indirekt die Störsignalanteile geschätzt. Eine Extraktion der Störsignalanteile aus den dem Spracherkennungssystem zugeführten akustischen Signalen kann somit überflüssig gemacht werden. Eine Schätzung der Störsignalanteile kann so erfolgen, dass vorgegebene akustische Referenzen in Abhängigkeit  
5 vom detektierten Betriebszustand und/oder von der detektierten Betriebsumgebung selektiert werden, um Sprachpausen zu modellieren, in denen die akustischen Signale lediglich Störsignalanteile aufweisen. Korrespondierend dazu lässt sich ein Vorliegen von Sprachsignalanteilen detektieren, was dann der Fall ist, wenn keine Sprachpause vorliegt; auf diese Weise kann ein fehlerhaftes Detektieren des Vorliegens von Sprachsignalanteilen  
10 bei der Änderung der Störsignalanteile vermieden werden. Durch die erfindungsgemäßen Maßnahmen wird die Zuverlässigkeit und Benutzungssicherheit des Gesamtsystems erhöht.

Auch Sprachsignalanteile repräsentierende akustische Referenzen lassen sich mittels der  
15 detektierten Betriebszustands- bzw. Betriebsumgebungsinformationen so anpassen, dass diesen überlagerte Störsignalanteile durch die akustischen Referenzen repräsentiert sind.

Eine zur Durchführung des erfindungsgemäßen Verfahrens geeignete Anordnung gibt  
20 Patentanspruch 8 an.

Für beliebige sprachgesteuerte Geräte wird die Aufgabe in entsprechender Weise gemäß den Merkmalen der Patentansprüche 8 (Verfahren) und 9 (Anordnung) gelöst.

Ausführungsbeispiele der Erfindung werden nachstehend anhand der Zeichnungen näher  
25 erläutert. Es zeigen:

- Fig.1 die wesentlichen Komponenten zur Durchführung des erfindungsgemäßen Verfahrens in einem Kraftfahrzeug,
- Fig.2 eine erste Möglichkeit zur Erzeugung einer akustischen Referenz für einen  
30 Sprachpausenabschnitt und
- Fig.3 eine zweite Möglichkeit zur Erzeugung einer akustischen Referenz für einen Sprachpausenabschnitt.

- Das in Fig.1 gezeigte Blockschaltbild beschreibt die Steuerung von Geräten oder Funktionseinheiten in einem Kraftfahrzeug. Die Geräte/Funktionseinheiten sind hier beispielhaft durch die Blöcke 1a und 1b dargestellt. Die Steuerung erfolgt mittels Sprach-
- 5 signalen, die über ein Mikrofon 2 einem automatischen Spracherkennungssystem 3 zugeführt werden, dessen Erkennungsergebnisse von einer Funktionseinheit 4 ausgewertet werden, die eine Umsetzung in an die Geräte/Funktionseinheiten 1a und 1b zu liefernde elektrische Steuersignale bewirkt.
- 10 Eine Funktionseinheit 5 kennzeichnet die Extraktion von Merkmalen von vom Mikrofon 2 gelieferten Mikrofonsignalen, wobei Merkmale für die einzelnen aufeinanderfolgenden Signalabschnitte üblicherweise zu Merkmalsvektoren zusammengefasst werden. Bei der Merkmalsanalyse wird ein akustisches Signal beispielsweise abgetastet, quantisiert und schließlich noch einer Cepstralanalyse unterzogen. Dabei erfolgt eine Aufteilung des
- 15 akustischen Signals in aufeinanderfolgende Rahmen, die sich teilweise überlappen; für jeden Rahmen wird ein Merkmalsvektor gebildet. Die Merkmalsvektorkomponenten werden durch die ermittelten Cepstralwerte gebildet. Funktionsblock 6 beschreibt übliche Vergleichprozeduren, bei denen die Merkmalsvektoren mittels üblicher Suchprozeduren mit einem akustischen Modell 7 verglichen werden, woraus sich das der Funktionseinheit
- 20 4 zugeführte Spracherkennungsergebnis ergibt. Dem Vergleich 6 und dem akustischen Modell 7 liegen sogenannte Hidden-Markov-Modelle zugrunde. Das akustische Modell 7 weist akustische Referenzen 8 und ein Lexikon 9 auf. Eine Wortuntereinheit von jeweils einem oder mehreren Phonemen ist dabei jeweils eine akustische Referenz zugeordnet. Durch das Lexikon 9 sind entsprechend den im Lexikon zusammengefassten Worten
- 25 zugehörige Folgen von Wortuntereinheiten definiert.

Das Spracherkennungssystem 3 weist eine Schnittstelle 10 auf, die eine Verbindung zu einem Bordcomputer 11 des Kraftfahrzeugs herstellt. Der Bordcomputer 11 wiederum ist durch eine Verbindung 12a mit mindestens einem Detektor 13 verbunden, der Betriebszu-

30 stands- und/oder Betriebsumgebungsinformationen detektiert und dem Bordcomputer 11 zuführt, der entsprechende Daten speichert. Die Betriebszustands- und/oder Betriebsumgebungsdaten werden der Schnittstelle 10 zugeführt, die diese Daten an eine Funktionsein-



- heit 12 weiterleitet, die zur Anpassung der akustischen Referenzen 8 an den jeweils detektierten Betriebszustand bzw. an die detektierte Betriebsumgebung anpasst. Grundsätzlich kann die Schnittstelle 10 auch ohne Zwischenschaltung eines Bordcomputers mit dem Detektor 13 gekoppelt sein (Verbindung 12b). Ein detektierter Betriebszustand wäre bspw. der Betriebszustand eines Lüfters oder auch die jeweilige Geschwindigkeit des Kraftfahrzeuges. Die Betriebsumgebungsdaten könnten bspw. Regenwetter indizieren oder auch den aktuellen Bodenbelag, auf dem das Kraftfahrzeug fährt.
- 10 Vorzugsweise können durch das beschriebene System Sprachpausenmodelle mit geeigneten akustischen Referenzen 8 erzeugt werden. Für Sprachpausen enthält ein vom Mikrofon 2 empfangenes akustisches Signal lediglich Störsignalanteile, jedoch keine Sprachsignalanteile, mit denen eine Steuerung der Geräte/Funktionseinheiten 1a oder 1b erfolgen soll.
- 15 Eine Ausgestaltung der Erfindung besteht darin, in Abhängigkeit von einem detektierten Betriebszustand oder einer detektierten Betriebsumgebung das Vokabular des Spracherkennungssystems 3, das durch das Lexikon 9 zusammengefasst ist, auf eine Untermenge von Worten einzuschränken, die als wirksame Sprachsteuersignale zur Verfügung stehen (Funktionsblock 13). Bei diesem Ansatz werden die für die VergleichsprozEDUREN des Funktionsblockes 6 erforderlichen Rechenoperationen reduziert.
- 20 Fig.2 zeigt ein Beispiel zur Erzeugung einer akustischen Referenz 8a aus der Menge der akustischen Referenzen 8. Mittels der Funktionseinheit 12 wird hier aus einer Menge a priori vorgegebener und vordefinierter Basisreferenzen 20-1, 20-2 bis 20-n für Sprachpausenabschnitte diejenige Basisreferenz ausgewählt, die schon vor der Inbetriebnahme des Spracherkennungssystems 3 demjenigen Betriebszustand bzw. derjenigen Betriebsumgebung zugeordnet wurde, und die dem aktuell detektierten Betriebszustand bzw. der aktuell detektierten Betriebsumgebung am besten entspricht. Die Auswahl einer Basisreferenz ist symbolisch durch einen Schalter 21 dargestellt. Funktionsblock 22 fasst eine optionale Adaption der selektierten Basisreferenz zusammen, um eine genauere Modellierung des detektierten Betriebszustandes bzw. der detektierten Betriebsumgebung zu erreichen und so die zu verwendende akustische Referenz 8a für den betreffenden
- 25
- 30

Sprachpausenabschnitt zu bilden. Wenn bspw. eine akustische Basisreferenz einem Störsignalanteil entspricht, der sich aus einem Regengeräusch ableitet, wird bei der Adaption gemäß Block 22 eine Anpassung an die detektierte Stärke des Regens erfolgen, wobei die Stärke des Regens mit einem entsprechenden Stör-/Rauschsignalpegel im Kraftfahrzeug korrespondiert.

Fig.3 zeigt eine weitere Variante zur Erzeugung der akustischen Referenz 8a für einen Sprachpausenabschnitt. Wie schon in Fig.2 sind a priori vorgegebene Basisreferenzen für Sprachpausenabschnitte (Blöcke 30-1, 30-2 bis 30-n) vorgesehen, mittels derer die akustische Referenz 8a gebildet wird. In der Ausführungsform gemäß Fig.3 wird allerdings nicht eine einzelne Basisreferenz selektiert. Es werden vielmehr alle Basisreferenzen einer Funktionseinheit 31 zugeführt, in der zunächst in Abhängigkeit von dem jeweils detektierten Betriebszustand bzw. der jeweils detektierten Betriebsumgebung eine Gewichtung und ggf. auch eine Adaption der Basisreferenzen durchgeführt wird (Blöcke 32-1, 32-2 bis 32-n). Die so gebildeten gewichteten/adaptierten Basisreferenzen werden schließlich in einer Einheit 33 zu einer einzigen akustischen Referenz kombiniert, welche die zu verwendende akustische Referenz 8a für den betrachteten Sprachpausenabschnitt ist.

Die Erfindung ist nicht auf die Sprachpausenmodellierung beschränkt. Grundsätzlich können auch die Wortuntereinheiten entsprechenden akustischen Referenzen 8 in entsprechender Weise an einen detektierten Betriebszustand bzw. eine detektierte Betriebsumgebung des Kraftfahrzeugs angepasst werden. Die akustische Referenz 8a würde dann die Grundlage für die Adaption von Wortunterabschnitten repräsentierenden akustischen Referenzen 8 bilden, um Störsignalanteile eines vom Mikrofon 2 aufgenommenen akustischen Signals zu modellieren.

Darüber hinaus ist die beschriebene Erfindung nicht auf den Einsatz in Kraftfahrzeugen beschränkt. Die Erfindung ist grundsätzlich anwendbar auf alle mittels Sprachsteuerung gesteuerten Geräte, bei denen Sprachsteuersignalen Störsignale überlagert sind, die sich indirekt durch Detektion des Betriebszustandes bzw. der Betriebsumgebung eines solchen Gerätes bestimmen lassen.

## PATENTANSPRÜCHE

1. Verfahren zum Steuern von Funktionseinheiten eines Kraftfahrzeugs oder von in einem Kraftfahrzeug angeordneten Geräten (1a, 1b) mittels Sprachsignalen, bei dem  
im Kraftfahrzeug auftretende akustische Signale, die vom Betriebszustand und/oder der Betriebsumgebung des Kraftfahrzeugs abhängige Störsignalanteile und gegebenenfalls Sprachsignalanteile enthalten, einem Spracherkennungssystem (3) zugeführt werden und  
das Spracherkennungssystem (3) akustische Referenzen (8) verwendet, die in Abhängigkeit von detektierten Betriebszustands- und/oder Betriebsumgebungsinformationen ausgewählt und/oder adaptiert werden.

10

2. Verfahren nach Anspruch 1,  
dadurch gekennzeichnet,  
dass akustische Basisreferenzen (20-1 ... 20-n, 30-1 ... 30-n) in Abhängigkeit vom Betriebszustand und/oder der Betriebsumgebung des Kraftfahrzeugs zur Verwendung für  
eine Sprachpausenmodellierung selektiert werden.

15

3. Verfahren nach Anspruch 2,  
dadurch gekennzeichnet,  
dass eine Adaption (22, 32-1 ... 32-n) der selektierten akustischen Basisreferenzen in  
Abhängigkeit vom Betriebszustand und/oder der Betriebsumgebung des Kraftfahrzeugs vorgesehen ist.

20

25

4. Verfahren nach Anspruch 1,  
dadurch gekennzeichnet,  
dass zur Sprachpausenmodellierung akustische Basisreferenzen in Abhängigkeit vom Betriebszustand und/oder der Betriebsumgebung des Kraftfahrzeugs miteinander  
5 kombiniert (33) werden.
5. Verfahren nach einem der Ansprüche 1 bis 4,  
dadurch gekennzeichnet,  
dass Betriebszustandsdaten und/oder Betriebsumgebungsdaten des Kraftfahrzeugs aus  
10 einem Bordcomputer (11) des Kraftfahrzeugs ausgelesen und/oder mittels ein oder mehrerer am Kraftfahrzeug montierter Detektoren (13) ermittelt werden.
6. Verfahren nach einem der Ansprüche 1 bis 5,  
dadurch gekennzeichnet,  
15 dass in Abhängigkeit von den detektierten Betriebszustands- und/oder Betriebsumgebungs-  
informationen des Kraftfahrzeugs diejenigen Teile des Vokabulars (9) des  
Spracherkennungssystems (3) bestimmt werden (13), die zum Steuern von  
Funktionseinheiten des Kraftfahrzeugs oder von im Kraftfahrzeug angeordneten Geräten  
(1a, 1b) wirksame Sprachsteuersignale darstellen.  
20
7. Anordnung zum Steuern von Funktionseinheiten eines Kraftfahrzeugs oder in einem  
Kraftfahrzeug angeordneter Geräte (1a, 1b) mittels Sprachsignalen mit  
- mindestens einem Mikrofon (2) zur Umwandlung von im Kraftfahrzeug auftretenden  
akustischen Signalen, die vom Betriebszustand und/oder der Betriebsumgebung des  
25 Kraftfahrzeugs abhängige Störsignalanteile und gegebenenfalls Sprachsignalanteile  
enthalten, in Mikrofonsignale und  
- einem mit dem Mikrofon (2) gekoppelten Spracherkennungssystem (3) zur Erkennung  
von Sprachsignalanteilen der akustischen Signale, wobei eine Auswahl und/oder ein  
Adaptieren von vom Spracherkennungssystem (3) verwendeten akustischen Referenzen  
30 (8) in Abhängigkeit von detektierten Betriebszustands- und/oder Betriebsumgebungs-  
informationen vorgesehen ist.

8. Verfahren zum Steuern eines Geräts mittels Sprachsignalen, bei dem
- akustische Signale, die vom Betriebszustand des Geräts und/oder der Betriebsumgebung des Geräts abhängige Störsignalanteile und gegebenenfalls Sprachsignalanteile enthalten, einem Spracherkennungssystem zugeführt werden und
- 5 - das Spracherkennungssystem akustische Referenzen verwendet, die in Abhängigkeit von detektierten Betriebszustands- und/oder Betriebsumgebungsinformationen des Geräts ausgewählt und/oder adaptiert werden.
9. Anordnung mit einem mittels Sprachsignalen steuerbaren Geräts, bei dem
- 10 - akustische Signale, die vom Betriebszustand des Geräts und/oder der Betriebsumgebung des Geräts abhängige Störsignalanteile und gegebenenfalls Sprachsignalanteile enthalten, einem Spracherkennungssystem zugeführt werden und
  - das Spracherkennungssystem akustische Referenzen verwendet, die in Abhängigkeit von detektierten Betriebszustands- und/oder Betriebsumgebungsinformationen des
- 15 - Geräts ausgewählt und/oder adaptiert werden.

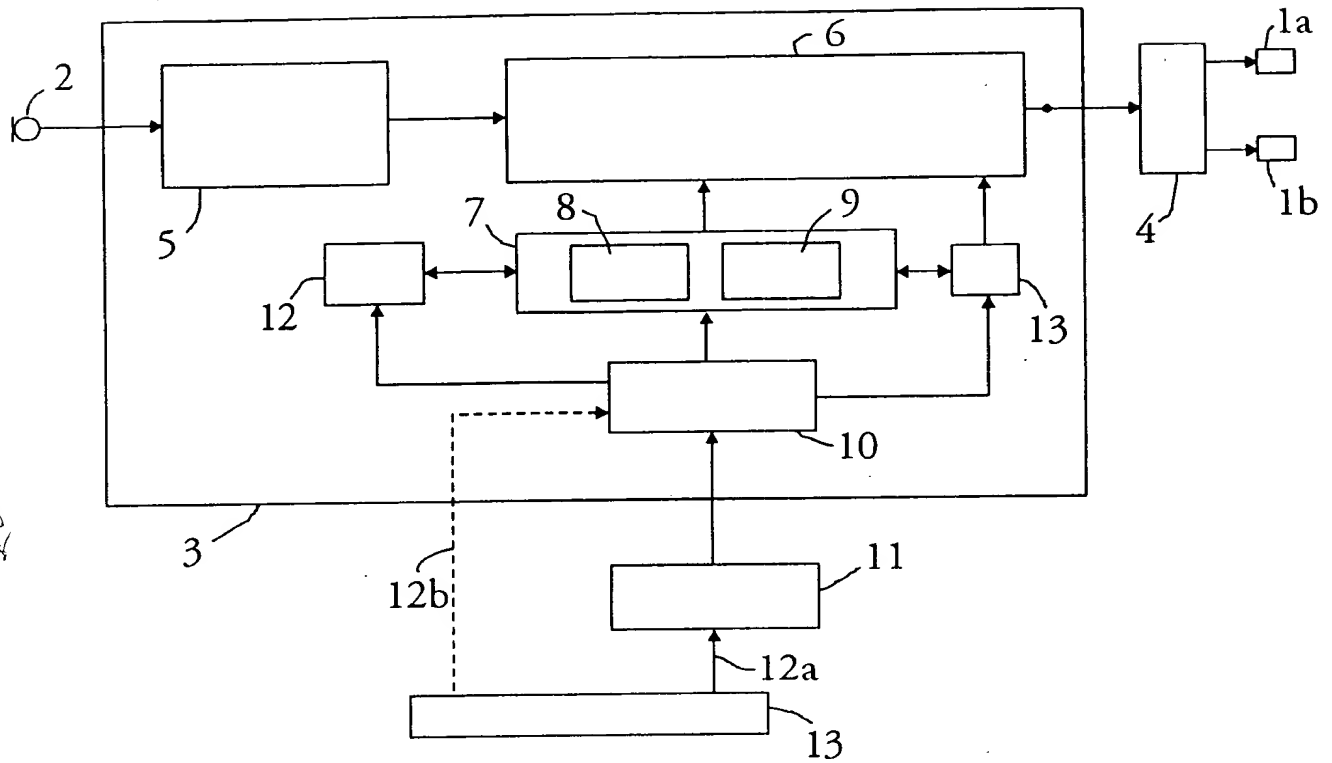


FIG. 1

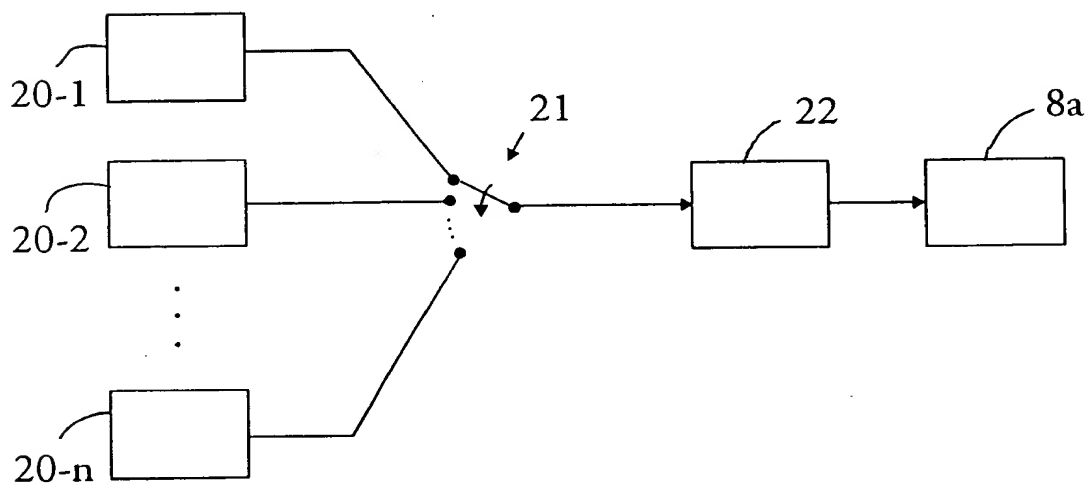


FIG. 2

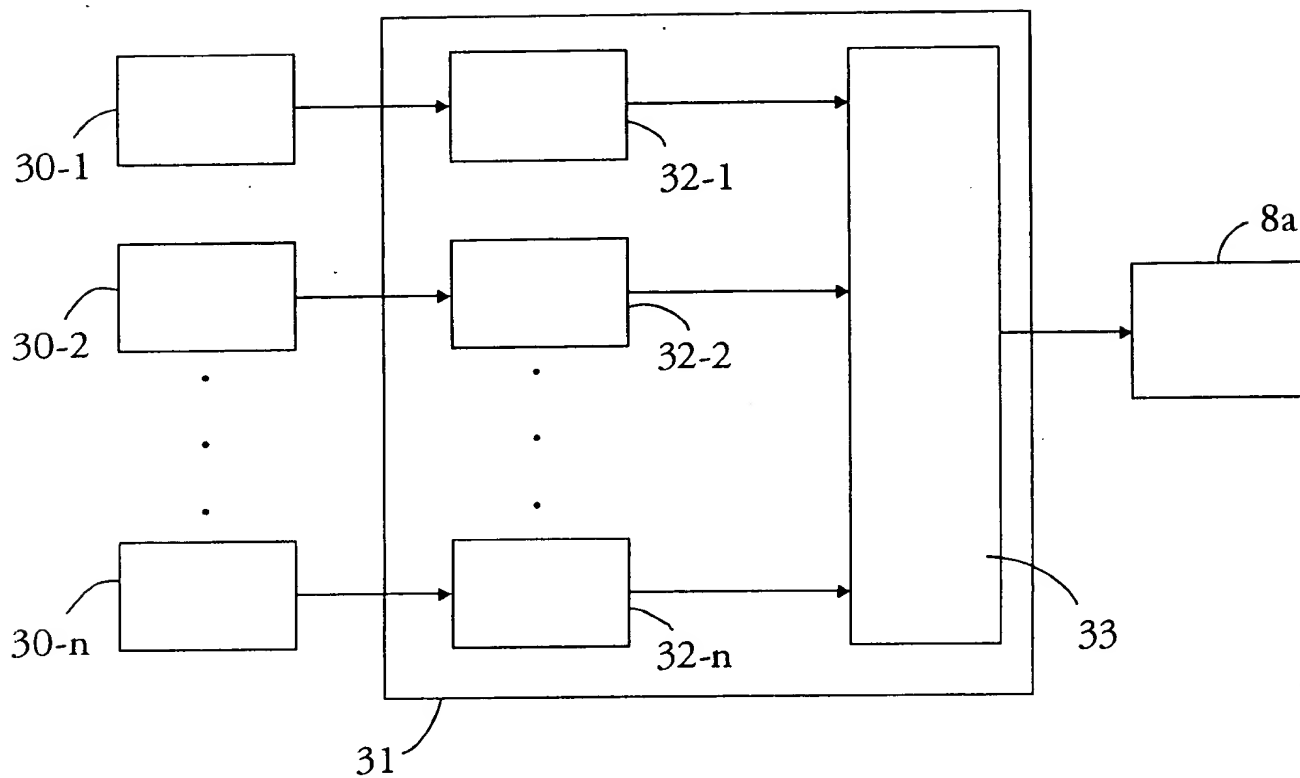


FIG. 3